Mapping and confirming new genes in *Arabidopsis thaliana* involved in formation of distinct cellular domains on pollen surface

Research Thesis

Presented in partial fulfillment of the requirements for graduation with research distinction in Molecular Genetics in the undergraduate colleges of The Ohio State University

by

Prativa Amom

The Ohio State University

April 2018

Project Advisor: Dr. Anna Dobritsa, Department of Molecular Genetics

Table of Contents

Abstract

Pollen wall exine is placed in species-specific patterns around pollen grains to protect them and facilitate plant reproduction. In the pollen of the model organism *Arabidopsis thaliana*, exine is deposited non-uniformly, always resulting in the formation of three longitudinal gaps not covered by exine. These gaps are called apertures, and they help pollen to control its moisture content in response to the environment and allow emergence of the pollen tube during pollen germination. The precision with which apertures are formed, and the fact that their patterns are diverse across species, make pollen apertures a powerful model for studying how cells specify and develop extracellular domains. Previously, only one gene, *INP1*, had been known to influence pollen aperture formation in *Arabidopsis.* In order to identify other genes involved in this process, a genetic screen was performed on mutagenized plants. Five complementation groups defective in aperture formation have been found, and positional cloning isolated gene candidates for four of these groups. To confirm the identity of two of these genes, for mutant groups *macaron* and *donut*, and to initiate their characterization, multiple constructs containing wild-type versions of these candidate genes were created with the addition of fluorescent tags and transformed into each mutant population. T_1 plants containing the constructs were selected and assessed for rescue of the wild-type phenotype.

Introduction

Pollen pattern formation is a finely controlled and conserved process, evidenced by the striking consistency in the pollen shape, size, and structures, across plants of the same species. The unique and structurally varied patterns that develop on the pollen surface (Fig. 1) are created through the deposition of exine, the highly durable and protective outer cell wall of pollen. Exine is composed of sporopollenin, a resilient compound whose structure is not yet fully understood but is known to be extremely chemically inert (Guilford et al., 1988). The development of exine is a great evolutionary advantage because it gives protection and allows plant sperm cells that reside inside pollen grains to survive detrimental environmental stresses like desiccation, UV exposure, and organic pathogens, significantly impacting the reproductive ability of plants (Edlund et al., 2004). Deposition and patterning of exine on the pollen surface is a highly regulated process that is exceedingly conserved within a given plant species. However, across species, exine deposition patterns vary vastly (Fig. 1). These unique exine patterns of different species are suspected to serve an important role by preventing plants from being pollinated by pollen of other species through a complex pollen recognition system (Zinkl et al., 1999). Exine deposition is usually non-uniform on the pollen surface, resulting in the formation of characteristic openings, the apertures (Fig. 1) Apertures are the sites where exine deposition is limited or absent, and they allow pollen grains to expand and contract along with the changing states of pollen hydration (Katifori et al., 2010). Apertures also frequently serve as exit points for pollen tubes carrying sperm cells during fertilization, thus playing an important role in male fertility and plant

reproduction (Edlund et al., 2004). The number, shape, size, and positions of apertures differ significantly across species, but like the exine patterns, aperture patterns are usually well conserved within a single species. Such precision in aperture formation indicates the existence of robust molecular mechanisms that protect specific domains on pollen surface from exine deposition. How these domains are specified and recognized by the molecular machinery involved in exine deposition is not yet understood. This precise patterning of pollen surface provides an excellent model to study the controlled process of deposition of extracellular materials, which in many cell types and in many organisms is involved in cell protection and regulation of cell growth, morphology, movement, and communication (Rozario and De Simone, 2010). Therefore, the exploration of this question will be applicable not only as a model for formation of extracellular domains in pollen grains but also in numerous other fields and systems.

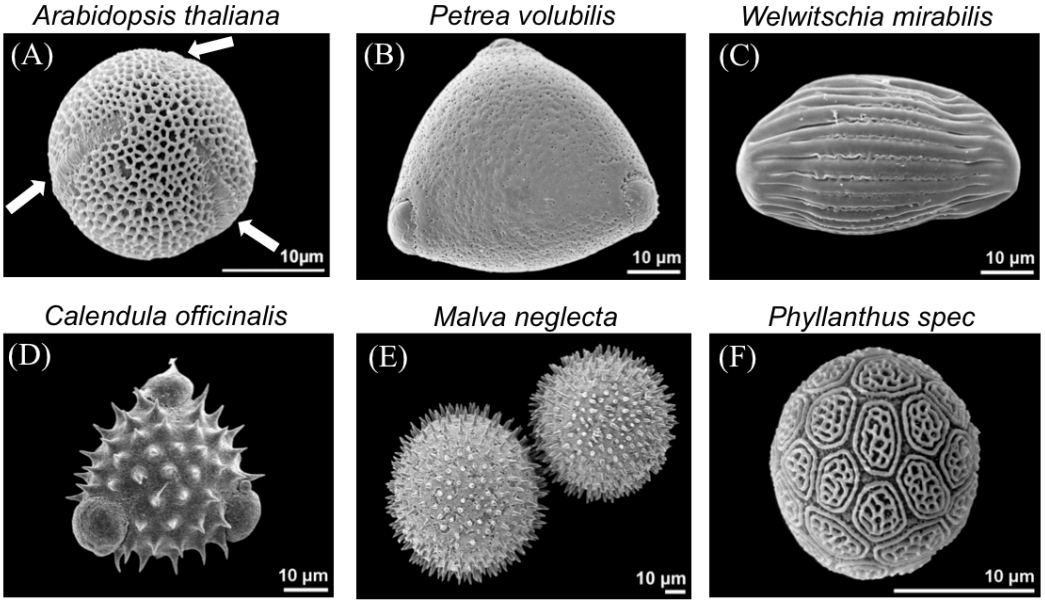


Figure 1
Scanning electron microscopy images of wild type pollen from PalDat database show a range of different exine patterns. Common names of the plants from which these pollen specimens were obtained include A) thale cress, B) queens wreath, C) tree tumbo, D) English marigold, E) common mallow, and F) gooseberry. Arrows point to apertures on the surface of Arabidopsis thaliana pollen.

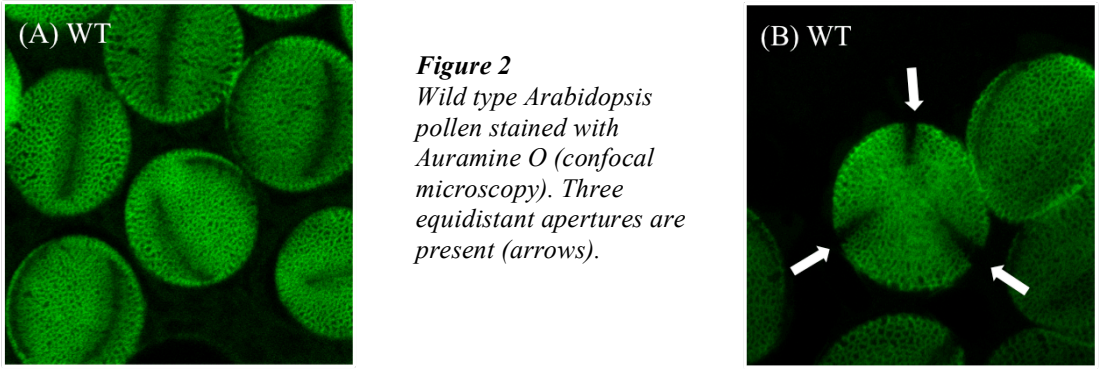


Figure 2
Wild type Arabidopsis pollen stained with Auramine O (confocal microscopy). Three equidistant apertures are present (arrows).

Background

In *Arabidopsis thaliana*, which we are using as a model system for studying exine patterns on pollen surface, the wild-type pollen grains have a reticulate exine pattern interrupted by three equidistant, longitudinal, and elongated apertures with the centers at the equator of the pollen grain (Fig. 1A, Fig. 2A). Pollen grain development begins inside the anther with a diploid microspore mother cell undergoing meiosis and creating four haploid microspores, initially grouped in a tetrad formation (Fig. 3C; Blackmore et al., 2007). Each of these four microspores will eventually develop into mature pollen grains.

Exine patterning is believed to initiate at the tetrad stage, when primexine, the precursor of exine, is deposited on the plasma membrane of the microspores (Ariizumi and Toriyama, 2011). The primexine deposition is followed by the deposition of sporopollenin, the constituent of exine that is transported from the tapetum, the inner layer of the anther, past the microspore mother cell wall and callose wall, onto the microspore surface. At the tetrad stage, the microspore plasma membrane starts undulating, and the highest points of undulation mark positions where columns of

sporopollenin are deposited next to the plasma membrane of the microspores, forming column-shaped baculae of exine (Ariizumi and Toriyama, 2011).

Aperture patterns also first become apparent in the tetrad-stage microspores as specific sites where primexine deposition does not take place and where plasma membrane undulation probably does not occur (Dobritsa et al., 2018).

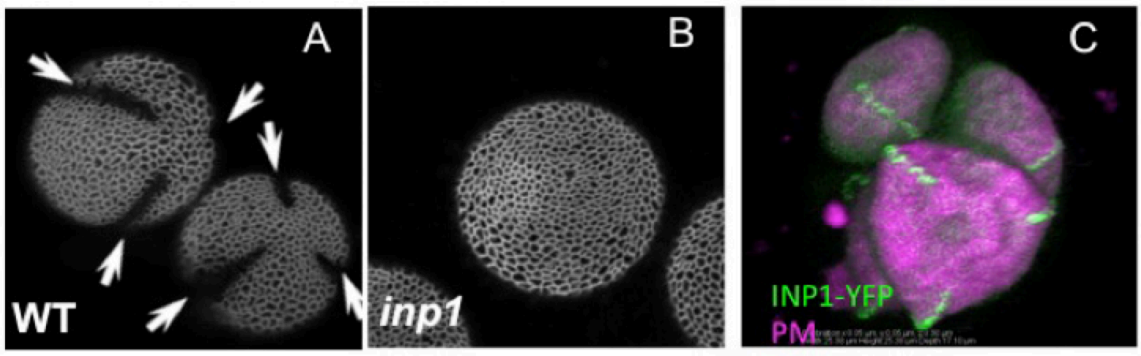


Figure 3
Shown here are confocal images of pollen stained with Auramine O: (A) Arabidopsis wild type (WT) pollen grains have three equidistant apertures (arrows). (B) inp1-1 mutants have a normal exine reticulate pattern but exhibit no apertures. (C) 3D reconstruction of confocal z-stacks of a tetrad of microspores in which INP1-YFP signal (green) forms three equidistant lines, resembling apertures, at the microspore periphery. Plasma membrane is stained here as magenta. Images used with permission of Dr. Anna Dobritsa.

Although more information on the steps of pollen development has been discovered recently, there is still very little known about how specific exine and aperture patterning is regulated (Shi et al., 2015). Until recently, only one *Arabidopsis* gene, *INP1*, encoding a novel protein, was known to influence pollen aperture formation. Previously found in an *Arabidopsis* mutagenesis screen, the null mutant alleles of this gene result in the development of round pollen that completely lacks apertures but has otherwise normally developed exine (Fig. 3B; Dobritsa et al., 2011, Dobritsa and Coerper, 2012).

Through further experimentation, the protein encoded by the *INP1* gene was found to localize specifically at the three equidistant longitudinal domains of the plasma membrane that correspond to the regions where apertures would later form on the surface of developing microspores (Fig. 3C; Dobritsa et al., 2012; Reeder et al., 2016; Dobritsa et al., 2018). This highly unusual protein localization pattern suggested that INP1 is used to premark the specific membrane domains destined to become apertures. It also suggested that developing pollen grains might have molecular mechanisms that guide INP1 to its distinct positions on the microspore surface.

The next steps would be to uncover other important genes that affect exine deposition and aperture formation and may either guide INP1 to its sites of localization or in other ways specify aperture domains on the microspore membrane surface. This is the area of research that I focused on in this study.

Research Direction I

Genetic Screen

In order to find more genes related to exine deposition and aperture formation, a large-scale screen utilizing ethyl methanesulfonate (EMS)-mutagenized *Arabidopsis* plants was conducted. The aim of this screen was to isolate mutations in genes, other than *INP1*, involved in the formation of patterns on pollen surface, with a particular focus on mutants that affect the formation of pollen apertures. EMS produces random point mutations via guanine alkylation, thereby causing guanine-cytosine base pairs to break and reform into adenine-thymine base pairs (Yongsig, et al., 2006)

This screen was done using mutagenized seeds of the Landsberg *erecta* (Ler) ecotype of *Arabidopsis thaliana*. An ecotype is a particular population of a species that is genetically distinct and occupies a particular habitat; in this case, the original plants were from Landsberg, Germany. Ecotypes are considered to be populations belonging to the same species, and in the case of *Arabidopsis*, many ecotypes have been well-studied, genotyped, and sequenced, including Ler.

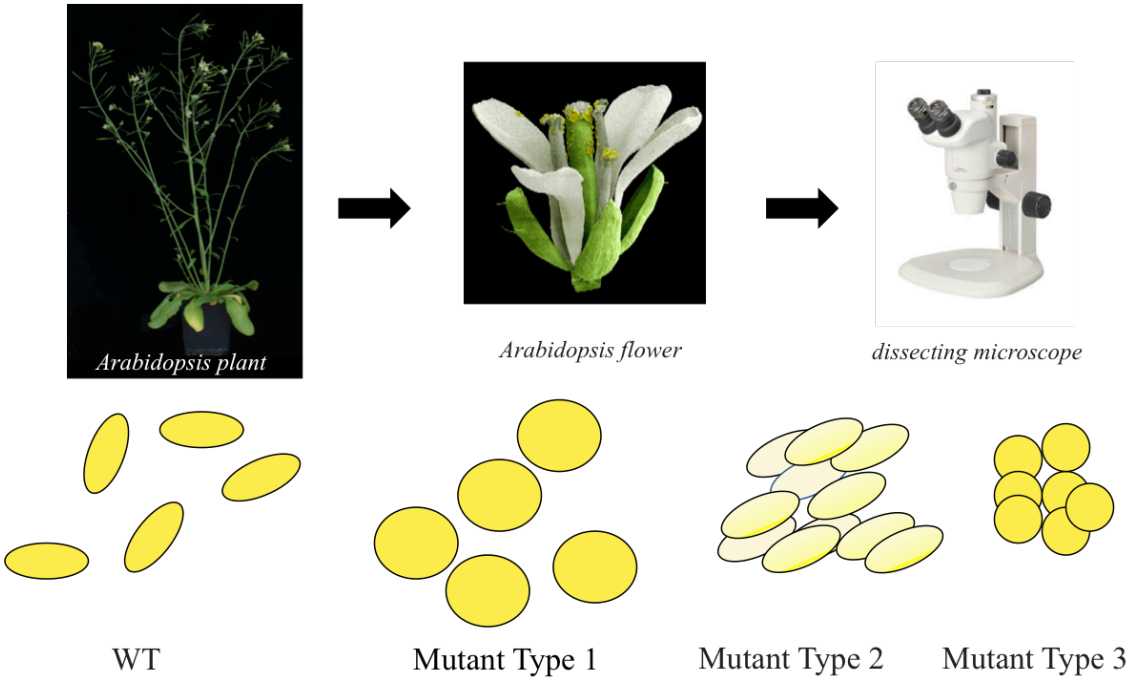


Figure 4
Illustration for the screening procedure done under a low-magnification dissecting microscope. Flowers were isolated from the Arabidopsis plant, and the anthers were dissected to remove pollen grains. Pollen was then scored as WT or mutant by looking at its color, formation of clumps, glossiness, shape, stickiness, etc… Not shown here, the next step was to stain pollen grains of mutant candidates and look at them at higher magnification under a confocal microscope.

Ler seeds were soaked in EMS, planted, and allowed to self-fertilize (M_0 generation). Most mutations in genes will be recessive, so in order for a mutant genotype to exhibit a mutant phenotype, two mutant alleles have to be present, creating a

homozygous condition. To facilitate the emergence of homozygous copies of these recessive alleles, the M_1 generation was planted and allowed to self-fertilize. In the M_2 generation, the phenotypic screen for pollen mutants was performed. For this, flowers from more than 11,000 individual plants were dissected. Anthers were removed, and pollen was screened for mutant exine and aperture phenotypes, initially at the level of low-magnification dissection microscopy (Fig. 4) as previously described (Dobritsa et al. 2011). Plants with pollen that looked glossy, sticky, especially clumpy, or white, or which did not have the normal oval shape were marked as potential mutants. Plants with pollen that was not clumpy/shiny, had a yellow sheen, and exhibited consistent grain-of-rice-like oval shape were considered to be wild-type and were discarded. The phenotypes of promising candidates observed underneath the dissecting microscope were then verified by high-magnification, high-resolution confocal microscopy.

Phenotype	Thin/No Exine	*spotty, spongy,* or *upex*-like	Large Lacunae	Fine Pattern with Small Lacunae	Abnormal Aperture	Larger Pollen with Many Apertures	Miscellaneous
Number of Isolates	29	111	5	14	22	12	303

Table 1 *Types of mutants with abnormal pollen surface and other pollen defects found in the EMS screen. A total of 11,014 individual plants were screened.*

A large number of potential mutants with defective pollen surface patterning were isolated, including nine mutants that had especially interesting defects in aperture formation (Table 1). Pair-wise complementation crosses between these nine mutants, as well as crosses with the *inp1* mutant, demonstrated that these new mutants affected genes different from *INP1* and that they fell into five new complementation groups. The

phenotypes of these mutants ranged from the complete loss of apertures (*inp1*-like phenotype) to changes in aperture number and positions (a single belt-like aperture instead of three apertures or two polar round apertures instead of three elongated equatorial furrows) to apertures that were shorter than normal or partially covered by exine (Fig. 5).

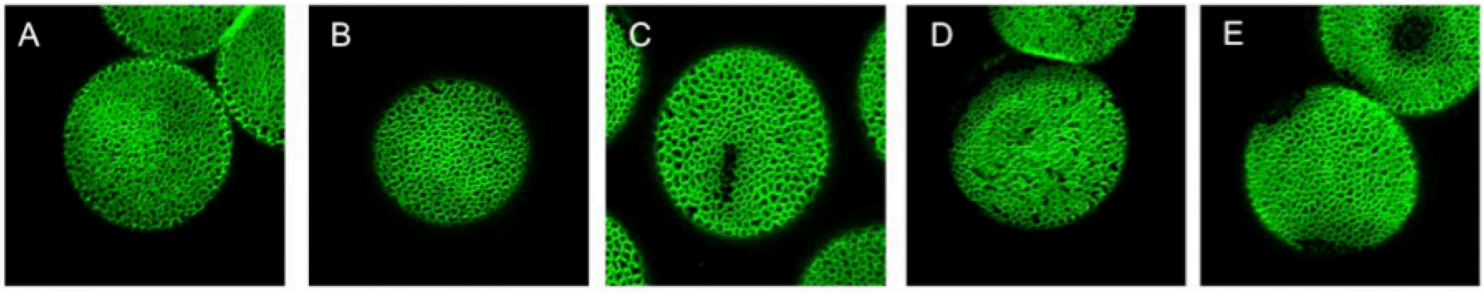


Figure 5
*Phenotypes of the newly discovered pollen aperture mutants range from a complete loss of apertures (**A** and **B**), shortened apertures (**C**), a single belt-like aperture (**D**), to two polar round apertures (**E**)*

Positional Cloning/Gene Mapping

In the next step, positional cloning techniques were applied to identify the specific genes affected in four of these complementation groups. This strategy failed with one complementation group due to a more variable phenotype that was difficult to reliably score at low magnification, resulting in the insufficient number of mutant plant samples for successful mapping. To map mutations, mutants with their background of Landsberg *erecta*, were crossed with wild-type plants from another Arabidopsis ecotype, Columbia (Col), originally from Columbia, Missouri (Fig. 6). The F_1 progeny of the cross between our mutant Ler and wild-type (WT) Col plants had one allele from each of their parents at the locus of mutation (heterozygous for the recessive mutant allele), and thus had the

phenotype of WT plants. This F_1 generation was then allowed to self-fertilize, allowing for recombination between the L*er* and Col chromosomes (Fig. 6). In the F_2 populations, approximately 25% of progeny exhibited mutant pollen phenotype and were expected to be homozygous at the locus of mutation and to have two copies of the mutant allele surrounded by the Ler chromosomal region (Fig. 6). Then, utilizing the polymorphisms between the L*er* and Col ecotypes, we set out to do the positional cloning, progressively narrowing the intervals containing the mutation loci.

This process required the screening of the F_2 generation to identify plants that had mutant phenotypes. Using a dissecting microscope, flowers of plants in the F_2 populations were dissected and their pollen was scored as either mutant or WT. The WT plants were discarded, and leaf tissue from each mutant plant was collected. For each mutant line of interest, the goal was to collect at least 300-400 samples of individual plant genomic DNA (gDNA) extracted from leaf tissue for positional cloning. A bulk segregation analysis (BSA) sample for each mutant line/complementation group was created by combining aliquots of hundreds of individual mutant plant DNA into one 'sample' for use in pool tests with different molecular markers in the next step.

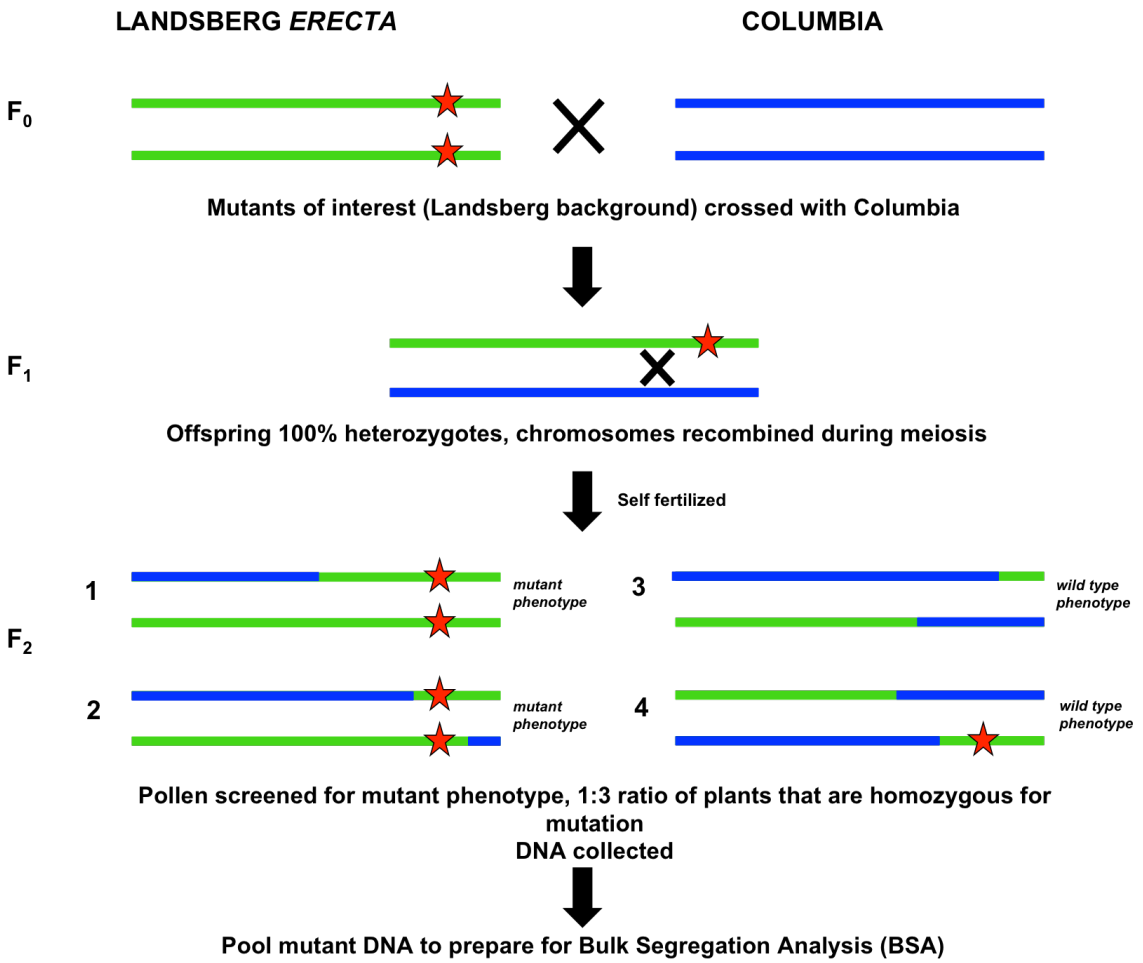


Figure 6

The population for positional cloning was set up ~~utilizing~~ by crossing mutants that had the Ler ecotype background with the WT Columbia ecotype. Bars represent homologous chromosomes. Red star represents position of a mutation. After another generation of self-fertilization, F2 plants were screened and DNA was extracted from 300-400 individual mutant plants. Aliquots of DNA from individual mutants was then compiled into a BSA sample. Both BSA samples and individual plant DNA samples were used for positional cloning applications.

After isolating 300-400 individual mutant gDNA samples of per mutant line, and preparing the BSA sample, we started isolating the locus of the mutation. *Arabidopsis thaliana* has five chromosomes covering about 132 mega-bases of DNA containing over 20,000 protein-coding genes (Woodward and Bartel, 2018). In order to systematically analyze such vast stretches of DNA, initial molecular markers that could distinguish between Col sequences and Landsberg sequences were created for several positions (5 to

8) along each of the five chromosomes to identify where in the mutant genomes homozygous Ler-specific marker(s) were enriched. For example, 1-2 markers for analysis of Ler/Col DNA were created near the top of each chromosome, several in the middle, and 1-2 closer to the end of the chromosome. These markers were chosen because they could distinguish between polymorphisms in Landsberg and Columbia genomic sequences when used with polymerase chain reaction reactions (PCR). A vast majority of these molecular markers were Simple Sequence Length Polymorphisms (SSLPs) or insertions/deletions (INDELs). However, some Derived Cleaved Amplified Polymorphic Sequences (dCAPS) markers were also used when SSLPs/INDELs that could differentiate between Ler and Col DNA were not accessible within a specific studied interval. These markers are different from SSLPs/INDELs due to their introduction of restriction enzyme (RE) sites during PCR. In the case of dCAPS markers, the PCR products later need to be digested in order to analyze the presence of either homozygous Ler, homozygous Col, or heterozygous alleles.

Molecular markers were first used on the BSA samples of each mutant line to determine the chromosome on which the aperture mutation likely resided by studying which of the markers consistently showed the presence of homozygous Ler sequence. On the other hand, the regions where much of the DNA was either homozygous for Col or heterozygous for Col and Ler could not be the locus of mutation. Once a large interval of homozygous Ler alleles was established, more molecular markers were selected within this interval and used to distinguish Ler and Col DNA in individual plant gDNA samples (Fig. 7). Since each mutant plant sample had different combinations of Col and Ler DNA in their genomes due to recombination but all were expected to have Ler DNA at the

mutant loci, the interval containing the mutation could be narrowed using information from individual plants. This was done using samples from individual mutant plants until the interval studied for each mutation was small enough that it contained approximately thirty to fifty genes and occupied approximately 150 kilo bases or less (Fig. 7).

Once the interval was small enough to look at individual genes, potential gene culprits were determined using programs provided by The Arabidopsis Information Resource (TAIR) database that predicted gene functionality and expression. Genes that were especially highly expressed in the appropriate floral development stages that aligned with the timing of pollen exine and aperture development were of special interest and were sequenced first as likely candidates. Other genes with predicted transmembrane domains, or genes involved in signaling or protein transport were also strong candidates due to the complex nature of membrane domain formation occurring through the deposition of proteins across membranes. Paying special attention to these characteristics, genes of interest that were postulated to be the most likely culprits were sent for sequencing using the Sanger sequencing procedure and then compared to the known WT sequence. Through this process, we were then able to map four genes from the five mutant lines.

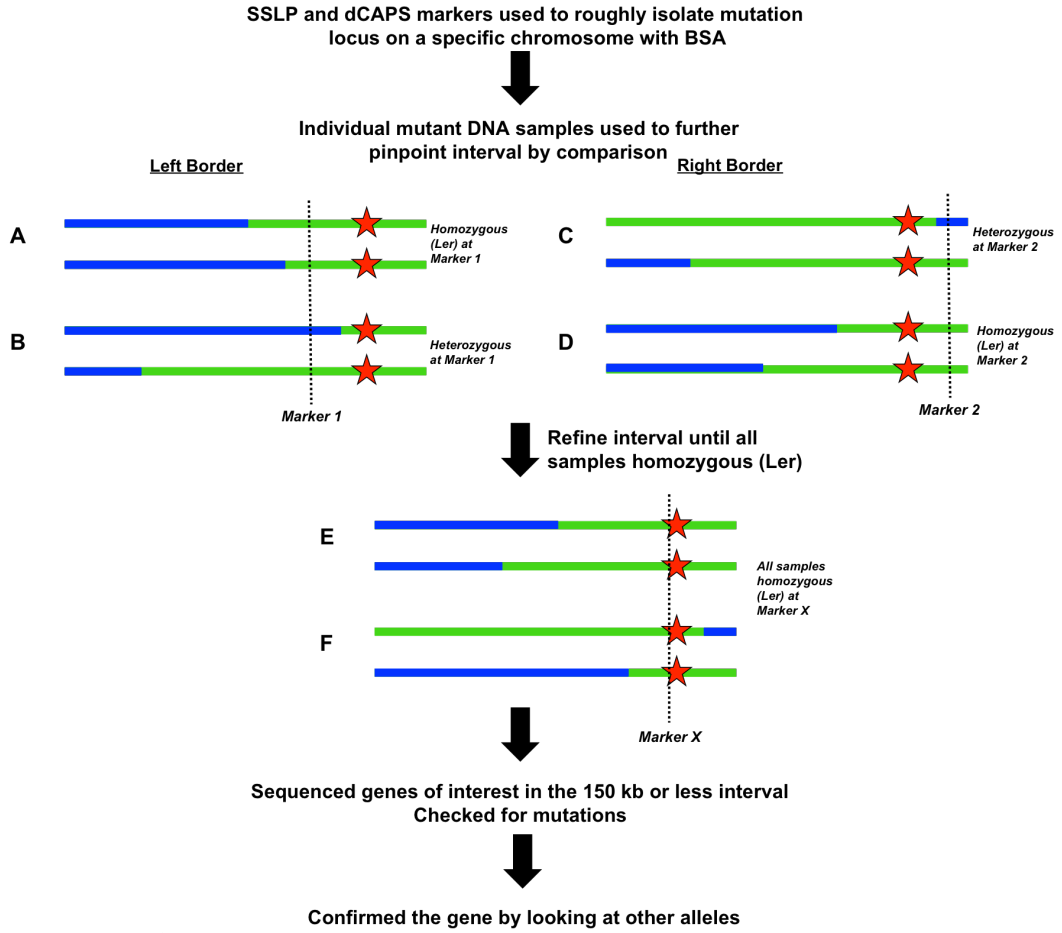


Figure 7

Mapping strategies used a range of markers to identify and narrow the DNA interval that could host the locus of mutation. Markers used include SSLPs, INDELs, and dCAPS. Once the interval was small enough, Sanger sequencing was used to look at individual genes.

The genes involved in pollen aperture formation identified in the screen

Four genes, potentially involved in pollen aperture formation, were identified using this approach: *MACARON (MCR,* encoding an *ELMO/CED-12* domain protein*), STRUBBELIG RECEPTOR FAMILY 2* (*SRF2*, a receptor-like kinase), and two novel genes*, INAPERTURATE POLLEN 2 (INP2)* and *DONUT (DNT)*. Mutations in the *MACARON* gene resulted in the *macaron* mutant phenotype, marked by a distinct single ring-like aperture (Fig. 5D, Fig. 8A). *SRF2* gene mutations caused *srf2* mutants to look aperture-less (Fig. 5A), similar to *inp2* mutants with mutations in the *INP2* gene (Fig. 5B) and to the previously discovered *inp1* mutants (Fig. 3B). *DONUT* mutations were also

very distinct because *dnt* mutants had two round apertures on the opposite sides from each other with some sporopollenin deposits inside (Fig. 5E, Fig. 8B). For the fifth complementation group containing the mutant with the short apertures (Fig. 5C), the discernment of mutant pollen via dissecting microscopes was very difficult. Because of this difficulty, there were only a limited number of mutant plant samples, and thus positional cloning data could not be sufficiently gathered to identify a culprit gene for this complementation group.

The locus of mutation causing the *macaron* mutant phenotype was identified as the At2g44770 gene. This gene encodes a protein of 366 amino acids (aa). Not much is known about the MCR protein, except that it belongs to the ELMO-CED12 family. The *mcr* mutants were found in the screen on four separate occasions, and complementation tests showed each original mutant plant had mutations in the same gene. In total, four different *mcr* mutant lines, containing mutations in different regions of the *MCR* gene, were gained from the screen. In addition, three more *mcr* mutant lines caused by transfer DNA (T-DNA) insertions were identified by searching the database of available T-DNA collections. Not much is known about the function of ELMO-CED12 family in plants. However, *Arabidopsis* has five additional homologues of this gene, while other plants also have orthologs of *MCR*.

For the *srf2* mutants, the locus of mutation was found to be the large gene At5g06820 that encodes a 735-aa protein. *srf2* mutants were found twice in the genetic screen. *SRF2* stands for *STRUBBELIG-RECEPTOR FAMILY 2*, and there are many representatives of the members of the *STRUBBELIG* families in many plants across species, with Arabidopsis having nine genes belonging to this family.

In the case of the similarly looking *inp2* mutants, the locus of mutation was found to be At1g15320, a smaller gene that encodes a novel protein of 307 aa. One mutation was found in the screen, and the protein domains and gene function remain unclear.

The gene disrupted in the *donut* mutant was identified as At2g37880, encoding another novel protein of 247 aa. Only one *dnt* mutant was found in the screen. The protein is known as MIZU-KUSSEI-like, but there is little else known about it. In the next part of my study, I focused on the genes *MACARON* and *DONUT* and tried to confirm the positional cloning results by proving that WT versions of these genes, when expressed in the *mcr* and *dnt* mutant plants, would rescue the mutant pollen phenotype.

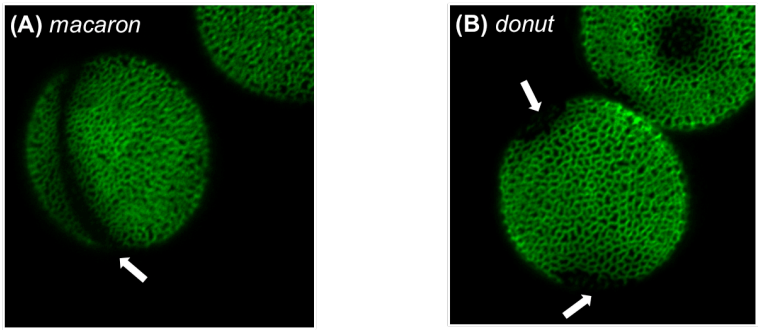


Figure 8 **(A)** *macaron mutants have one belt-like aperture.* **(B)** *donut mutants have two polar round apertures.*

Research Direction II

Gene Cloning

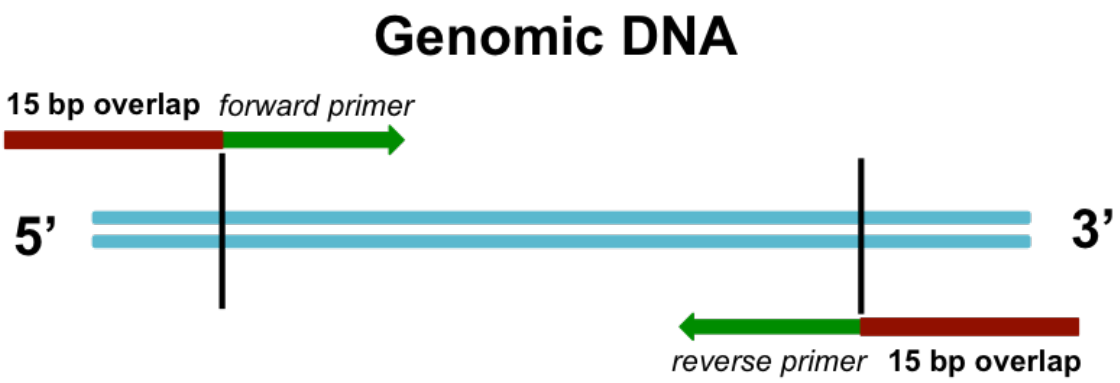


Figure 9
The In-Fusion approach that was utilized for cloning and creation of transgenic complementation constructs used fragments with 15-bp-long overlapping ends. Promoter and open reading frame fragments were cloned from genomic DNA or cDNA using high fidelity DNA polymerase Phusion with constructed primers that had 15 bp ends overlapping with the vector insertion area.

To confirm these gene identification results, in the next step, we transgenically expressed WT copies of the genes in mutant plants to observe if the phenotypes would be rescued. My project specifically focused on the genes *MACARON* and *DONUT*.

To create transgenic complementation constructs, the *MCR* and *DNT* genes (either cDNA or genomic versions) were placed under the control of their endogenous promoter regions or under a promoter from a different microspore-expressed gene, *DMC1*. The cDNA versions of the genes of interest were obtained from the constructs available from the Arabidopsis Biological Resource Center (ABRC). Since cDNA is constructed from mRNA, it contains only spliced exons of the gene of interest, and this allows for easier cloning of shorter DNA fragments. The open reading frame (ORF) of the gene was amplified for cloning using cDNA as a template. This ORF approach often works well in

Arabidopsis. Genomic DNA (gDNA) was copied from the whole unedited gene and included introns and exons, as well as parts of the putative 5' and 3' untranslated regions (UTRs).

Multiple approaches for cloning of promoters, ORF DNA, and gDNA were utilized. Special high-fidelity Phusion DNA polymerase was used to amplify the genes of interest using PCR. Another technique was the use of In-Fusion primers in PCR. In-Fusion primers add 15 base pairs on each end of the DNA segment of interest, overlapping the area where it would later insert into the plasmid vector (Fig. 9). Because the genome of *Arabidopsis* is known, these In-Fusion primers were easily designed. Using these primers, promoter regions, ORF DNA, and gDNA were amplified via PCR.

Construct Making

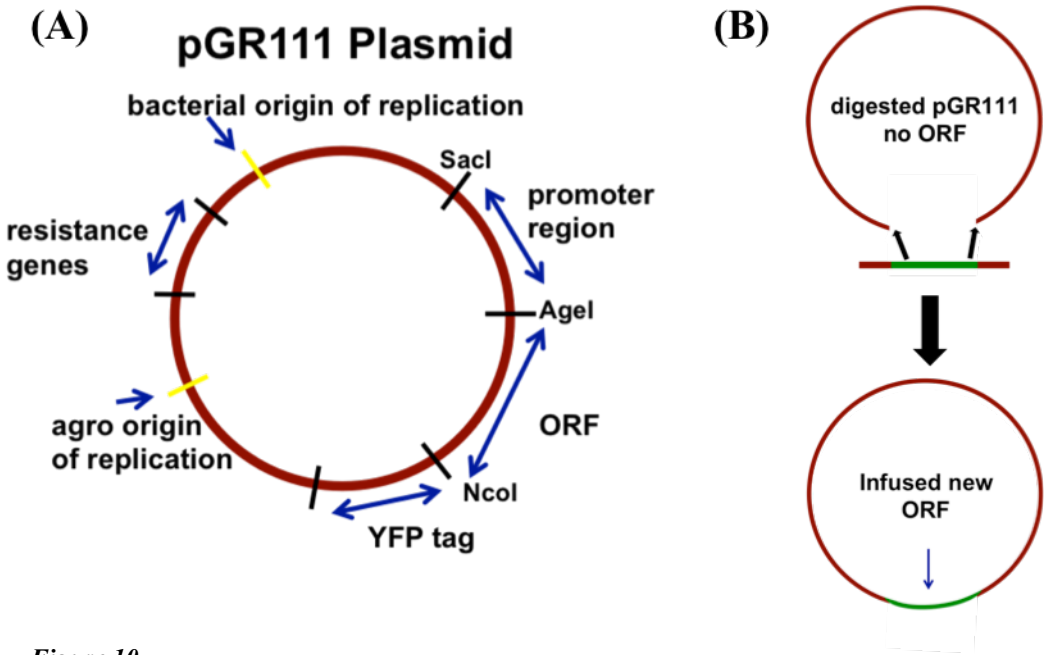


Figure 10
(A) *The backbone pGR111 plasmid was cut using restriction enzymes SacI and NcoI to put in promoter and ORF or genomic DNA.* ***(B)*** *PCR-amplified promoter and ORF/gDNA were inserted via In-Fusion method into digested plasmid using 15 bp overlapping sticky ends + In-Fusion enzyme.*

Both In-Fusion cloning technique (Clontech) and ligation after restriction enzyme digestion were utilized in the creation of transgenic constructs. The In-Fusion technique is especially easy to use: the base vector construct is linearized with specific restriction enzymes at the sites where the DNA segment of interest will be inserted. The addition of the In-Fusion enzyme and an incubation period of 15 minutes allows for the recombination of the insert DNA into the construct via the 15 bp overlap left by the primer. This approach allowed cloning of multiple fragments (e.g. a promoter and an ORF) simultaneously into the vector. A less straightforward way to create constructs was to cut both construct and DNA insert with the same pair of restriction enzymes and use ligation to allow the DNA insert to get incorporated into the vector. This method has a lower success rate, but is also cheaper.

Initially, the ORFs for genes of interest were inserted into a plasmid vector, pGR111, containing either a *DMC1* promoter (the promoter active in the microspore mother cell and microspores, which worked well in previous experiments with *INP1* (Reeder et al., 2016; Dobritsa et al., 2018)) or an endogenous promoter for *MCR* or *DNT*. These constructs also included a Yellow Fluorescent Protein (YFP) tag, which was fused to the gene of interest at the C-terminus to allow visualization of the corresponding protein. Later, we also created untagged ORF or genomic versions for both genes, as well as a version of DNT that was tagged with a derivative of YFP, mCitrine, at the N-terminus. Also included in the pGR111 vector was a gene providing resistance to the herbicide BASTA, which was used for selecting plants carrying the transgene (Fig. 10). Another gene in the pGR111 plasmid encoded resistance to the antibiotic kanamycin (Fig. 10) and was used for selecting *E. coli* bacteria that carried the constructs of interest.

Specific built-in restriction sites were also included in the construct to allow for easy insertion of DNA of interest into the plasmid, while the presence of bacterial origins of replication allowed for propagation of these constructs in *E. coli* and *Agrobacterium* (Fig. 10). Finally, the construct contained the left and right border of T-DNA from Agrobacterium to allow these sequences to get transferred and incorporated into the plant genome.

These constructs were then transformed into competent *E. coli* cells, plated, and colonies were selected for resistance to kanamycin. Liquid cultures of healthy colonies were prepared and used to amplify and isolate plasmid constructs. A restriction enzyme digest was then performed on the isolated plasmids: DNA fragments separated on an agarose gel via electrophoresis were analyzed to verify the correct fragment sizes of the construct. After this verification, the constructs were sent for sequencing to ensure that no mutations were introduced during the PCR amplification. The constructs were then transformed into *Agrobacterium*.

Transformation

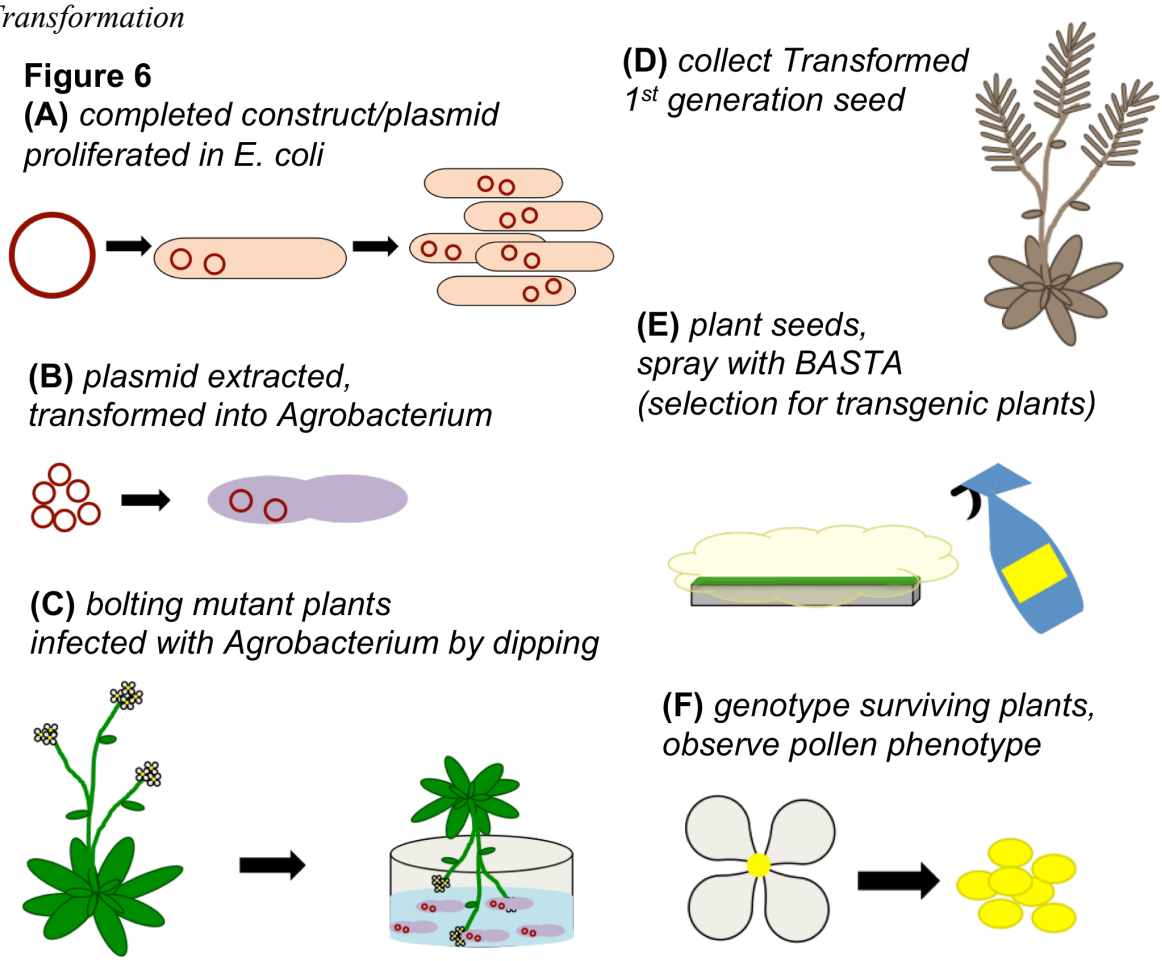


Figure 11
The workflow used for transformation of transgenes into Arabidopsis thaliana mutant lines.

The workflow used for creation of transgenic plants is shown in Fig. 11. Confirmed constructs were transformed into electro-competent GV3101 *Agrobacterium* cells: colonies containing the construct had resistance to kanamycin. Liquid cultures of healthy individual *Agrobacterium* colonies were grown and mixed with detergent and sucrose. This solution was then used to infect the mutant line *Arabidopsis* plants using the floral-dip method (Clough and Bent, 1998). Taking advantage of *Agrobacterium*'s ability to infect plants with the DNA plasmids, bolting plants from each mutant line were

dipped into cultures of Agrobacterium containing constructs with WT versions of their mutant genes (Fig. 11). After these plants matured, their seeds were collected and dried. Next, the seeds of the infected plants, the T_1 generation, were planted. An herbicide called BASTA was then sprayed on seedlings to select for and identify plants containing transgenic constructs (Fig. 11). These plants were also genotyped to confirm the presence of transgenic constructs. The pollen phenotypes of the transgenic plants were then observed under the dissection microscope. The pollen shape, visible at this low-magnification level, can serve as a good proxy for the aperture phenotype – with WT pollen having oval shape, *mcr* pollen having many triangular grains, and *dnt* pollen exhibiting round shape. Thus, at this point, we could usually predict if the pollen of transgenic plants still continued to exhibit a mutant phenotype or were successfully complemented. After that, the pollen samples were then prepared for confocal microscopy to confirm the aperture phenotypes and to see if expression of the WT gene can rescue the mutants.

Complementation of mcr and dnt mutant phenotypes

Initially, constructs containing the *DMC1* promoter and ORF copies of *DNT* and *MCR* tagged with YFP at the C-terminus were created. Both constructs failed to rescue the mutant pollen phenotype when transformed into the corresponding mutant plants.

The next batch of constructs which I made used the endogenous promoter of the gene of interest instead of *DMC1*. These two constructs were *MCR*pr:*MCR* ORF-*YFP* and *DNTpr*:*DNT* ORF-*YFP*. When mutant lines transformed with these two constructs were observed under the dissecting microscope, *dnt* transformants did not exhibit any

sign of rescue of the mutant phenotype. On the other hand, *mcr* mutants transformed with *MCRpr:MCR* ORF-YFP exhibited partial rescue of the WT phenotype.

A third batch of constructs was then created to test for the interference of YFP with protein functionality and for the effect of the regions that were absent in the previous ORF constructs (Fig. 12). Two more *MCR* constructs were created. One was *MCRpr:MCR* ORF and the other was *MCRpr:MCR* gDNA, both without a fluorescent tag. Even under the dissecting microscope, it was clear that the *mcr* mutant plants transformed with both of these constructs exhibited strong rescue of the WT phenotype. Three more constructs of *DNT* were created in the same batch (Fig. 12). The first two were followed a similar pattern to the *MCR* constructs created. They were the untagged *DNTpr:DNT* ORF and *DNTpr:DNT* genomic DNA. Another *DNT* construct was made with the addition of the Citrine tag at the N-terminus of the protein (Fig. 12). The plants transformed with the untagged gDNA construct exhibited clear rescue. However, neither the *DNTpr:DNT* ORF nor *DNTpr:Citrine-DNT* ORF construct were able to rescue the mutant phenotype in the *dnt* plants (Fig. 12).

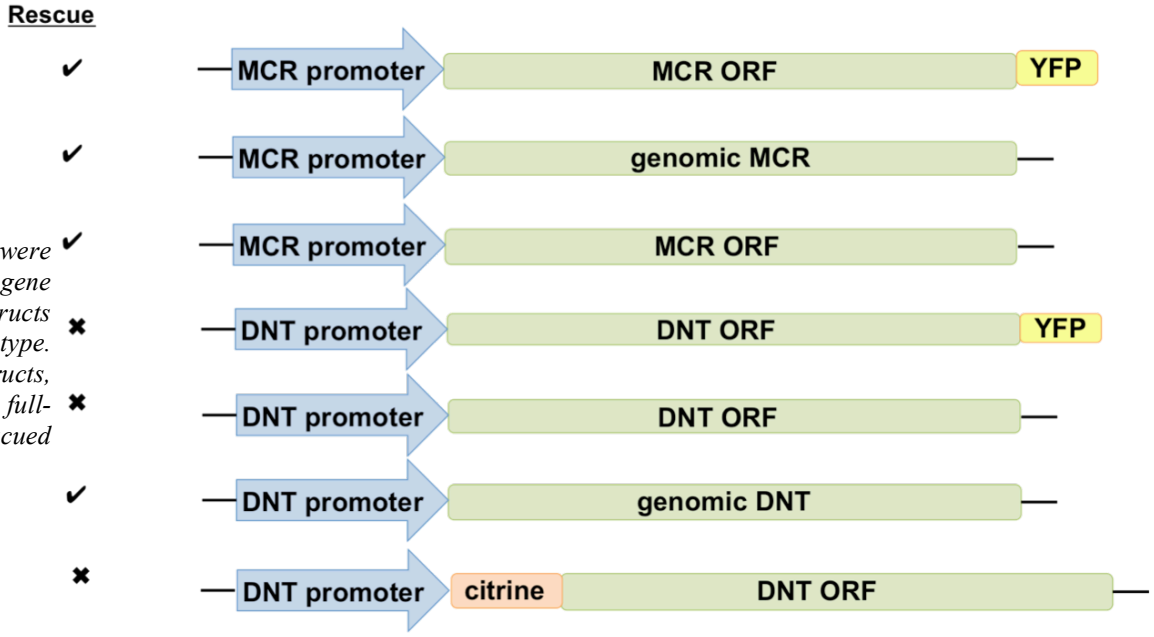


Figure 12
Seven different constructs were created with endogenous gene promoters. Three MCR constructs all rescued the mutant phenotype. Out of the four DNT constructs, only one, containing the full-length genomic sequence, rescued the mutant phenotype.

Discussion

In this study, we performed a genetic screen that successfully identified five new aperture complementation groups. Using positional cloning, we were able to map the four mutations to specific genes.

The results of the transformation of both *MACARON* and *DONUT* constructs suggest that since the strong rescue of mutant lines was possible through transformation of WT versions of *MCR* and *DNT* genes, they are indeed the culprit genes whose mutant forms cause the *macaron* and *donut* mutant phenotype. These results are further verified by the discovery of multiple alleles for both of these genes either in our screen or in available collections of annotated T-DNA insertions. However, not all of the complementation constructs worked equally well. Constructs with the *DMC1* promoter did not rescue at all, suggesting that their endogenous promoters are especially important and expression of these genes must be specifically regulated to perform its intended functions. As *DMC1* promoter is expressed in microspore mother cells, its failure to function in our experiments signals that both *MACAROON* and *DONUT* may have very small and specific time windows for gene activation for proper protein expression.

A few other constructs were also not successful in rescuing pollen phenotype, possibly because certain structures might have interfered with proper protein folding. One potential culprit is the YFP tag attached to the end of the C-termini in these genes. The YFP protein is 238 aa-long, which is comparable in size to both MCR (266 aa) and DNT (247 aa). So doubling the size of these proteins through addition of YFP might have interfered with folding of proteins and their ability to function. In the case of DNT, we

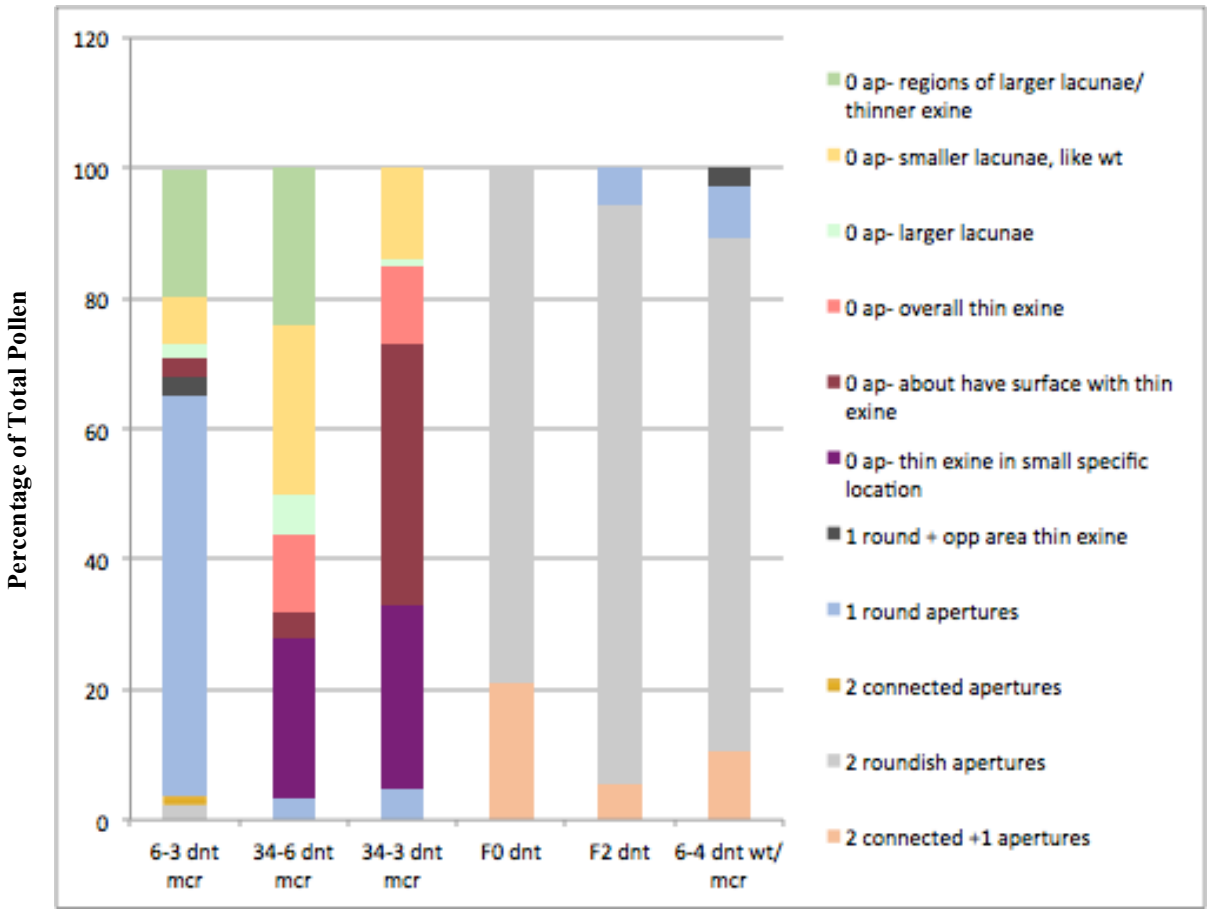
also tried to add the modified YFP (Citrine) to the N-terminus of the protein, however, this construct also failed to rescue.

For *MACARON*, the *MCRpr*:*MCR* ORF-*YFP* allowed partial rescue compared to the constructs without tags. Again, this was probably due to the YFP interfering with proper protein folding. However, when the *DNT* ORF was used, even without any tags, no rescue was observed. At the same time, the full-length copy of the genomic *DNT* demonstrated strong rescue. The main difference between the ORF and the genomic DNA constructs was the presence of 5' and 3' UTRs and introns. These results indicate that that the presence of UTRs or introns is important for proper *DNT* expression: these elements may, for example, control RNA stability or protein translation.

The next question to ask is what happens when a double mutant of *macaron* and *donut* is created. When plants homozygous for *DNT* were crossed with homozygous *MCR* pollen, double *donut* and *macaron* heterozygotes were created. *MCR* and *DNT* are both located on chromosome 2, not very far (~2.6 Mb) from each other: therefore, chromosomal recombination is required to produce double mutants and is expected to be fairly infrequent. The F_1 generation was allowed to self-fertilize, and the plants exhibiting the *dnt* phenotype in the F_2 generation were sequenced to identify *MCR* heterozygotes, which would be indicative of the presence of one chromosome containing mutations in both genes. After that, these F_2 *dnt/dnt mcr/+* plants were allowed to self-fertilize, and 12 plants in the F_3 generation were sequenced. Three plants were found to be homozygous for both *mcr* and *dnt* mutant alleles. Pollen from these double mutants was observed with confocal microscopy (Fig. 13). Although the aperture phenotypes in the pollen of double mutants were somewhat variable, the double mutants clearly exhibited a pattern distinct

from the single mutants. In all three double mutants, the number of apertures was usually reduced, with many pollen grains completely lacking apertures or having just one aperture. When apertures were present, they had round morphology, like apertures in the *dnt* mutant and very different from the more normal, narrow and elongated apertures of *mcr*. These results suggested that the two mutations had an additive phenotype, and that the *dnt* mutation is epistatic to the *mcr* mutation.

In addition, somewhat unexpectedly, some pollen also exhibited exine-related phenotypes, ranging from very large exine lacunae to extremely thin exine. It is possible that *MACARON* and *DONUT* are not the only players here and that recombination or specific combinations of chromosomes in these double mutants introduced an allele of another gene in some mutants but not the others. This range of phenotypes hints that there could be more genes that are playing roles in exine and aperture formation and have complex relationships with each other either in linear pathways or in positive or negative feedback loops.



Aperture and exine phenotypes in dnt mcr double mutant plants and in dnt or dnt mcr/+ plants

Figure 13
Graphically represented results of the phenotypes observed in the progeny of a cross between dnt and mcr. A number of different phenotypes were observed. For each individual line, the percentage of the total population that have a specific phenotype are shown. Sample sizes for each line are over 200 individual pollen grains. Study and counting were done under a confocal microscope.

Materials and Methods

Screening EMS-mutagenized Arabidopsis population:

The M_2 population of EMS-mutagenized *Arabidopsis thaliana* (in Landsberg *erecta* background) was planted and grown until flowering. Each plant was screened under the dissecting microscope for abnormal pollen phenotypes as described (Dobritsa et al., 2011). Pollen of promising candidates was stained with auramine O as described (Reeder

et al., 2016) and observed under the confocal microscope. Plants with wild-type-looking pollen were discarded.

Generation of mapping populations:

Pollen mutants with the Landsberg *erecta* background were crossed with the Columbia ecotype. The F_1 progeny were allowed to self-cross. Their progeny, the F_2 population, exhibited segregation of WT to mutant plants at an expected 3:1 ratio. All plants were grown in growth chambers under conditions of 18 hours of light and 6 hours of dark. The F_2 plants were screened under a dissecting microscope to identify segregating mutants. From each mutant plant, tissue samples were collected for genomic DNA preparation.

Isolation of genomic DNA:

For individual tissue samples, leaves were separately hand-grinded in Eppendorf tubes using a plastic pestle. When dealing with larger sets of tissue samples, liquid nitrogen was used to flash-freeze samples in Eppendorf tubes in bulk. They were then immediately lyophilized for the next 24 hours until desiccated. In order to grind these samples, 2.3 mm diameter zirconia-silica beads were added to the tubes and samples were vortexed until a fine green powder evenly stained the walls of the tubes. For both protocols of grinding tissue, Dr. Raha's buffer was then added. A 100 mL stock of this buffer consisted of 10mL of 10% SDS, 2mL of Triton-X 100, 2mL of 5M NaCl, 500µL of 2M Tris-HCl at a pH of 8.0, and 200µL of 0.5M EDTA. Samples were then mixed by hand and an equal volume of phenol:chloroform:isoamyl alcohol solution (25:24:1 ratio) was added to each sample. The samples were then vortexed for approximately 30 seconds and then

centrifuged for 10 minutes at 8,000 rpm, resulting in the separation of two phases. 80µL of the top layer was transferred to labeled Eppendorf tubes. These DNA-containing samples were diluted in water in a 1:5 ratio for use in PCR, cloning, and other experiments.

Use of molecular markers for mapping the mutations:

For each 25µL PCR reaction, a solution of 17.8 µL of H_2O, 5µL of MyTaq-5x buffer including dNTPs (Bioline), 0.5µL each of the 10 µM forward and reverse primer (constructed molecular marker), 0.2µL of MyTaq DNA polymerase (Bioline), and 1µL of 1:5 gDNA were used. Using a regular PCR cycler, samples were amplified through the 35 cycles of denaturation at 98°C for 15 seconds, annealing with temperature dependent on primer sequence and length, and elongation at 72°C, with time of elongation dependent on fragment length. Controls of WT Col, WT Ler, and heterozygous DNA were also run, along with mutant samples, for ease of comparison. For dCAPS markers, restriction digestion of PCR products was necessary with the correct restriction enzyme. PCR products were combined with 3µL of loading dye and run in a 1.5% agarose gel stained with ethidium bromide at 120 V with 5µL of a 100bp marker for 30-45 minutes. These gels were then imaged under UV light, and comparisons between controls and samples were performed to determine which samples were homozygous Ler, Col or heterozygous at that specific region of DNA.

DNA digestion:

For DNA digestion, a total 20μL restriction reaction was prepared. Each reaction consisted of 10μL of the PCR product, 2μL of the correct 10x New England Biolabs buffer dependent on the particular restriction enzyme, and 0.4μL of the corresponding New England Biolabs restriction enzyme. The reaction was then incubated at 37°C for about an hour. For both the 20μL digest product and the 25μL of regular PCR product, 3μL of loading dye is added. The samples were then run out via gel electrophoresis with an appropriate agarose gel concentration based on the predicted fragment size. A 100 bp or 1kb marker (New England Biolabs) was also run on the same gel to correlate band size.

Phusion high-fidelity PCR:

50μL total PCR reactions were run for the specific goal of cloning parts of genomic DNA and cDNA to create transgenic constructs. These reactions contained 28 μL of H_2O, 10μL of Phusion-5x buffer (New England Biolabs), 4μL of 2.5mM dNTPs, 2.5μL each of the 10 μM forward and reverse primer (constructed molecular marker), 0.5 μL of Phusion DNA polymerase (New England Biolabs), 1.5 μL of dimethyl sulfoxide (DMSO), and 1μL of 1:5 gDNA. Using a regular PCR cycler, samples were amplified through the 35 cycles of denaturation at 98°C for 15 seconds, annealing with temperature dependent on primer sequence and length, and elongation at 72°C, with time of elongation dependent on fragment length. The PCR products were checked for the presence of target fragments via gel electrophoresis.

Ligation:

DNA fragments to be ligated (e.g. a vector and an insert) were digested with the same set of restrictions enzymes, and correct fragments were gel-purified. For ligation, a total of 10 µL mixture was prepared, consisting of 1 µL 10x ligation buffer (New England Biolabs), 2 µL H_2O, 1µL T4 ligase (New England Biolabs), and equal concentrations of both fragments. Ligase was added last, and the mixture was incubated at room temperature for 2 hours.

In-Fusion cloning reactions:

Fragments to be cloned with the In-Fusion procedure (Clontech) had 15 bp overlapping ends and were generated through PCR. 1 µL of diluted promoter/genomic DNA/ORF DNA of interest, with 15 bp overlapping ends, was combined with similar concentration of 1 µL of the restriction enzyme-digested vector, 6 µL of H_2O, and 2 µL of In-Fusion Mix (Clontech). This mixture was then incubated at 50°C for 15 minutes for the DNA fragments to recombine and become incorporated into circular DNA (plasmid).

Sequencing:

In order to prepare PCR samples for sequencing, each fragment of the gene of interest was amplified in 50µL PCR reactions. A mixture of 35.6µL of H_2O, 10µL of MyTaq buffer with dNTPs included, with 1µL each of the 10µM forward and reverse primer, plus 0.4µL of the MyTaq DNA polymerase, and finally 2µL of the 1:5 diluted genomic DNA was made. PCR was done following the same conditions as described before. Afterwards, 5µL of PCR product was ran on a gel to check that amplification was

successful. For cleaning the PCR product to prep for sequencing, 5µL of PCR product was aliquoted and incubated with 0.2µL of Exonuclease I and 0.4µL of Shrimp Alkaline Phosphatase (SAP) in a 10µL reaction. This reaction was incubated at 37°C for 20 minutes and then at 80°C for another 15 minutes. 2.5µL of the purified PCR product was then mixed with 1µL of 10µM primer of interest in one direction and 2.5µL of dH_2O. The sample was then sent for Sanger sequencing analysis to a local facility.

Transformation to E. coli:

To transform constructs into *E. coli*, premeasured 50 µL aliquots of competent DH5α *E. coli* cells were thawed on ice. After thawing, 2 µL of the construct was gently mixed with the *E. coli* cells and incubated on ice for 30 minutes. After this, cells were heat shocked in a water bath at 42°C for 30 seconds and placed on ice for another 5 minutes. One mL of LB media was added to the tube and gently mixed and incubated at 37°C with shaking between 165-275 rpm for 1 hour. With this final incubation, the tube was centrifuged to form a pellet of bacteria at the bottom. The top 800 µL of the supernatant was decanted and the rest used to re-suspend the pellet and was plated onto LB agar plates with kanamycin added. Plates were incubated at 37°C overnight and checked for bacterial growth the next day. If colonies were present, they were checked for the presence of the correct construct by colony PCR or by plasmid purification and digestion.

Plasmid Miniprep:

To extract plasmids from *E. coli*, liquid 5 mL LB cultures containing 5 µL of kanamycin were inoculated with the transgenic *E. coli* colony of interest and left to grow overnight at

37°C on a shaker. The next day, 1.5 mL of the culture was loaded into an Eppendorf tube and centrifuged at 11,000 rcf for 30 seconds. The supernatant was carefully decanted and the pellet was saved. This step was repeated twice. The plasmid miniprep kit (NucleoSpin Plasmid, Macherey-Nagel) was used to extract DNA from bacterial cells and collect the plasmid of interest.

Transformation to Agrobacterium:

50 µL of pre-aliquoted frozen competent *Agrobacterium tumefaciens* (strain GV3101) was thawed on ice. 1 µL of the mini-prepped plasmid of interest was then added to Eppendorf tube containing agrobacteria with 1 µL of helper plasmid pSoup and gently mixed. This mixture was transferred to a chilled cuvette and electroporated at 1440 V. The cuvette was then placed on ice immediately and 1 mL of LB media was gently pipetted into cuvette. The liquid was transferred to a culture tube grown at 28°C with shaking for 1.5 to 2 hours. 200 µL of this solution was plated on LB plates containing 34 µg/mL Rifampicin, 50 µg/mL kanamycin, and 50 µg/mL of gentamicin. Plates required two days at 28°C for growth, and afterwards individual colonies were picked, added to 5 mL of LB media containing of 10 µL of 17 mg/mL Rifampicin, 5 µL of 50 mg/mL kanamycin, and 5 µL of of 50 mg/mL gentamicin and grown overnight at 28°C with shaking. Culture PCR was used to test for the presence of the target construct/plasmid.

Transformation to Arabidopsis thaliana:

2 mL of cultures of agrobacteria grown overnight were added to 500 mL of LB media containing 34 µg/mL Rifampicin, 50 µg/mL kanamycin, and 50 µg/mL gentamicin. This

larger culture was also grown overnight with shaking at 28°C The next day, the 500 mL of agrobacterium culture were separated into two large centrifuge bottles and centrifuged to pellet bacterial cells. The supernatant was decanted and cells were re-suspended in a solution of 400 mL of distilled H_2O, 20 g of sucrose, and 200 μL of Silwet L-77 (Helena). Mutant plant lines that were at the stage of bolting flowers were dipped in this detergent/agrobacterium mixture for 30 seconds. After dipping, the plants were allowed to produce seeds. Seeds were collected, planted in soil and transformants containing plasmids/constructs of interest were selected by spraying with the herbicide-BASTA.

Acknowledgements

I would like to give special thanks to Michelle Tan and Dr. Anna Dobritsa for their cooperation and direction on this project. This daunting task would not have been possible without the partnership of Michelle and all her hard work. Dr. Dobritsa's continued support and encouragements throughout the tenure of this project were invaluable. The National Science foundation and the Mayers Undergraduate Research Fellowship graciously provided financial assistance. Thank you also to my numerous role models and colleagues in the Dobritsa lab for technical help and good advice.

References:

Ariizumi, Tohru, and Kinya Toriyama. Genetic regulation of sporopollenin synthesis and pollen exine development." Annual review of plant biology 62 (2011): 437-460.

Ariizumi, T., and K. Toriyama. "Pollen exine pattern formation is dependent on three major developmental processes in *Arabidopsis thaliana*." Int J Plant Dev Biol 1 (2007): 106- 115.

Blackmore, S., Wortley, A.H., Skvarla, J.J., and Rowley, J.R. (2007). Pollen wall development in flowering plants. New Phytol. 174 483–498.

Clough, Steven J., and Andrew F. Bent. Floral Dip: a Simplified Method for Agrobacterium-Mediated Transformation of *Arabidopsis thaliana*. *The Plant Journal*, vol. 16, no. 6, 1998, pp. 735–743.

Dobritsa, A. A., Kirkpatrick, A., Reeder, S. H., Li, P. and H. A. Owen. Pollen aperture factor INP1 acts late in aperture formation by excluding specific membrane domains from exine deposition. *Plant Physiology*. Focus Issue on Plant Cell Dynamics, 2018, 176: 326-339.

Dobritsa, A. A., & Coerper, D. (2012). The Novel Plant Protein INAPERTURATE POLLEN1 Marks Distinct Cellular Domains and Controls Formation of Apertures in the *Arabidopsis* Pollen Exine. *The Plant Cell*, *24*(11), 4452–4464.

Dobritsa, A. A., Geanconteri, A., Shrestha, J., Carlson, A., Kooyers, N., Coerper, D., … Preuss, D. (2011). A large-scale genetic screen in Arabidopsis to identify genes involved in pollen exine production. *Plant Physiology*, *157*(2), 947–970.

Edlund, A. F., Swanson, R., & Preuss, D. (2004). Pollen and stigma structure and f unction: the role of diversity in pollination. *The Plant Cell*, *16*(Suppl), S84–S97.

Guilford, W. J., Schneider, D. M., Labovitz, J., & Opella, S. J. (1988). High Resolution Solid State ^{13}C NMR Spectroscopy of Sporopollenins from Different Plant Taxa. *Plant Physiology*, *86*(1), 134–136.

Katifori, E., Alben, S., Cerda, E., Nelson, D. R., & Dumais, J. (2010). Foldable structures and the natural design of pollen grains. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(17), 7635–7639.

Kim, Yongsig, et al. "EMS Mutagenesis of Arabidopsis." *Arabidopsis Protocols*, 2006, pp. 101–104.

Rozario, T. and DeSimone, DW. The extracellular matrix in development and morphogenesis: A dynamic view. Dev Biol. 2010; 341: 126–140.

Reeder, Sarah H., et al. "A Ploidy-Sensitive Mechanism Regulates Aperture Formation on the Arabidopsis Pollen Surface and Guides Localization of the Aperture Factor INP1." *PLOS Genetics*, vol. 12, no. 5, 2016.

Shi, J., Cui, M., Yanh, L., Kim, Y., Zhang, D. (2015). Genetic and biochemical mechanisms of pollen wall development. *Trends in Plant Science*, *20*(11), 741-753.

Woodward, A. and Bartel, B. (2018). *Biology in bloom: A primer on the Arabidopsis thaliana model system.* Genetics Society of America

Zinkl, G., Zwiebel, B., Grier, D., Preuss, D. Pollen-stigma adhesion in Arabidopsis: a species-specific interaction mediated by lipophilic molecules in the pollen exine. *Development.* 1999 Dec; 126 (23):5431–5440.